A Solution for the Implementation for
FixYourMoney.com
May 2022
Version: 1.0

Company: FixYourMoney.com
Industry Focus: Virtual debit card provider that adds credit card features
Established: 1st April 2020

Version His

No.	Date	Comments
1.0	16-May-2022	Outlined the structure for the Solution
1.1	30-May-2022	Solutioning the business case

Disclaimer: *The company, FixYourMoney.com, is a fictitious company used as an example to provide analysis towards the study of AI Technologies as part of the degree course for M.Sc. in Artificial Technologies from the University of Essex.*

1 Table of Contents

2 Introduction

Banking and finance are now better equipped with data-mining applications to identify financial risks earlier and more predictably.

Recommendations and credit scoring engines have become critical decision making during cross-selling, providing offers and lending money (Moin, K.I., & Ahmed, Q.B. (2012).

In this solution overview, the criticality of banks to make a better decision for the loan approval process has been considered. The behavioural aspects of the borrowers and their history are derived through the quality of data, the demographic profile, and the characteristics of the attributes (Hamid, Aboobyda & Ahmed, Tarig. (2016)).

3 Business Case

In the following sections, the business case of "credit risk" has been evaluated using the CRISP-DM methodology of data mining and testing through algorithms.

Focus has been put primarily on the challenge of refining the data to reduce the time consumed in dealing with large data sets.

The business case highlights the techniques and approach implemented in creating the model using a sample data set for the prediction.

The business case is built on providing a prediction based on the customer's historical record on whether the customer can be classified as good or bad to be given a loan.

4 Data Set

A collated data set using the below references of the credit risk sample data was used

- https://www.kaggle.com/datasets/uciml/german-credit
- https://www.kaggle.com/datasets/vipin20/loan-application-data
- https://www.kaggle.com/datasets/krantiswalke/bank-personal-loan-modelling
- https://archive.ics.uci.edu/ml/datasets/Bank+Marketing (Moro et al., 2014)
- https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients (Yeh, I. C., & Lien, C. H. (2009)

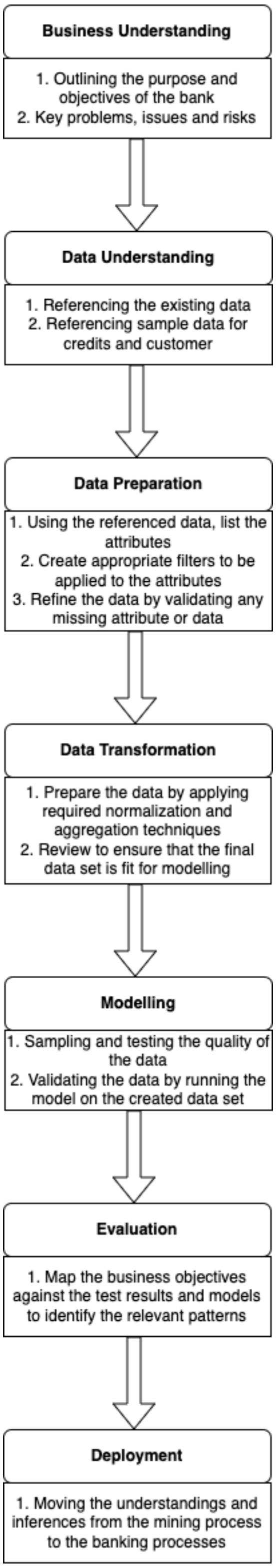
After reviewing the above references, the csv was formatted to be refined with the following attributes in the data set.

Attributes	Description	Data Type
Credit_history	Previous history of customer credit	Nominal
Existing_credits	Number of existing credits at this bank	Numeric

Purpose	The loan purpose	Nominal
Financed_amount	The amount of credit	Numeric
Gender	Male or Female	Nominal
Age	Age of the borrower	Numeric
Job	If the borrower has a job	Nominal
Class	The class of loan if good/bad	Nominal

5 Model preparation (CRISP-DM Analysis)

To ensure that the business case is reflected with appropriate predictions, the data-mining process involved using the CRISP-DM methodology. Below is the flow showing the steps to build the relevant model using the data set.



The attributes and values for the data set (above) were then refined and standardised using this methodology the below:

<u>Credit history attribute values</u>
'all paid'
'existing critical credit.'
'delayed previously' 'existing paid/not yet due.'

<u>Gender attribute values</u>
male
female

<u>Job attribute values</u>
qualified / self employed
skilled
student
unemployed / retired
unskilled

<u>Class</u>
good (Credit record status is “Current”, indicating loan payments have been paid accordingly)
bad (Credit record status is “Past-Due Non-Performing / Past-Due Not Yet Non-Performing / Items In Litigation.” displaying inconsistent payment records)

6 WEKA application

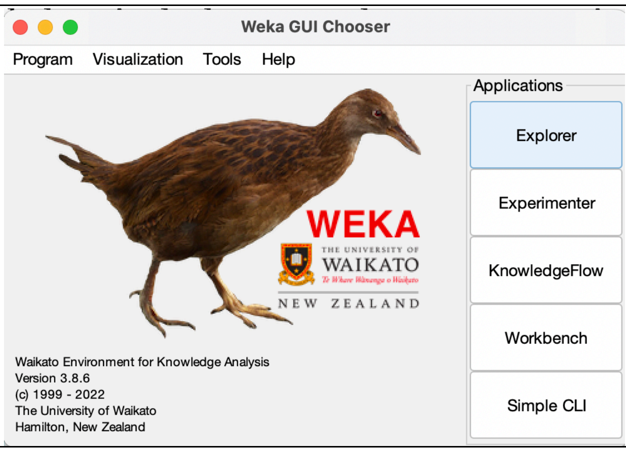
After a thorough CRISP analysis of the data, the data was then refined into a CSV format to be applied to the WEKA application tool by converting it into an appropriate ARFF format.

The most important attributes that were identified were

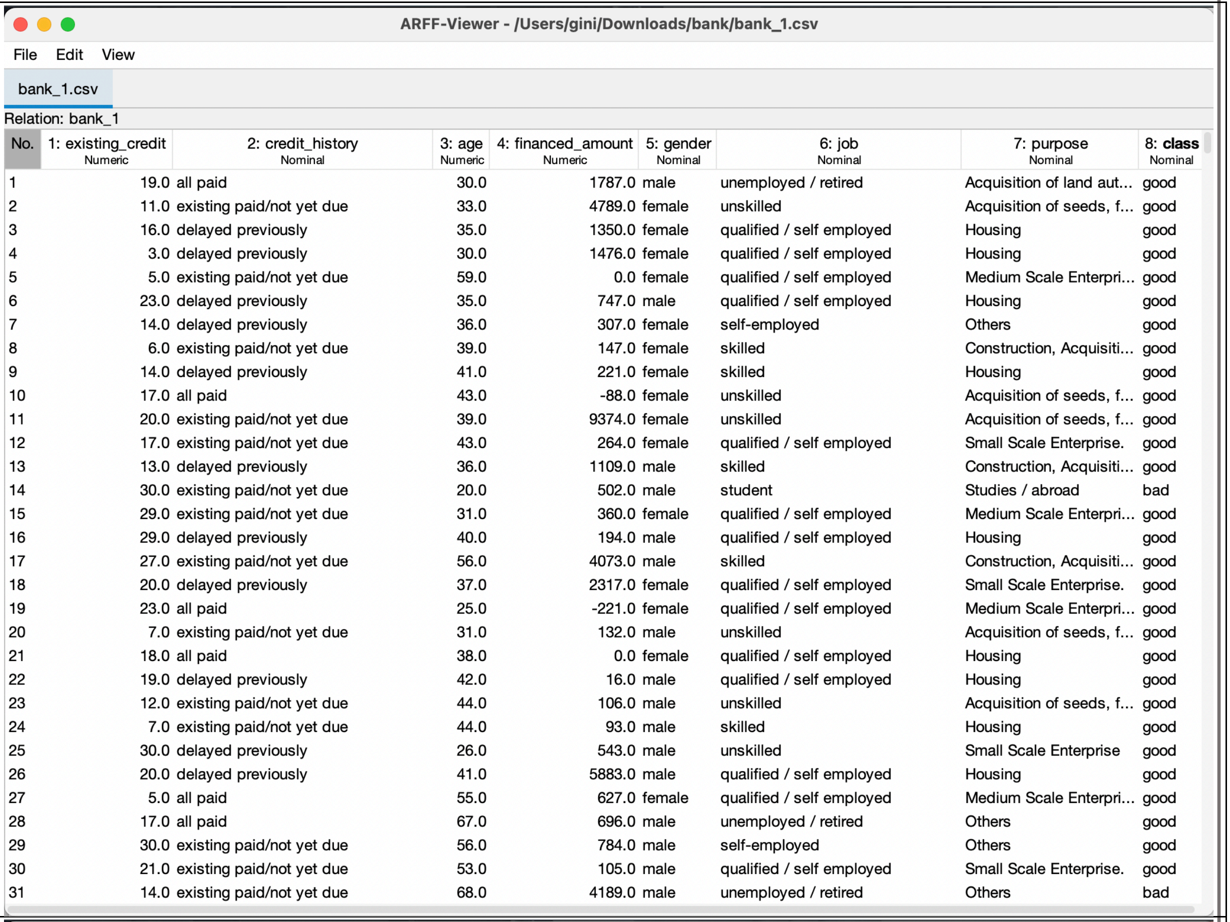
- Credit_history
- Existing_credits
- Purpose
- Financed_amount
- Gender
- Age
- Job
- Class: Good / Bad

(Hamid, Aboobyda & Ahmed, Tarig. (2016).)

The data is then fed into the WEKA tool for pre-processing.

WEKA 3.8.6	

Converting CSV to arff



ARFF-Viewer - /Users/gini/Downloads/bank/bank_1.csv

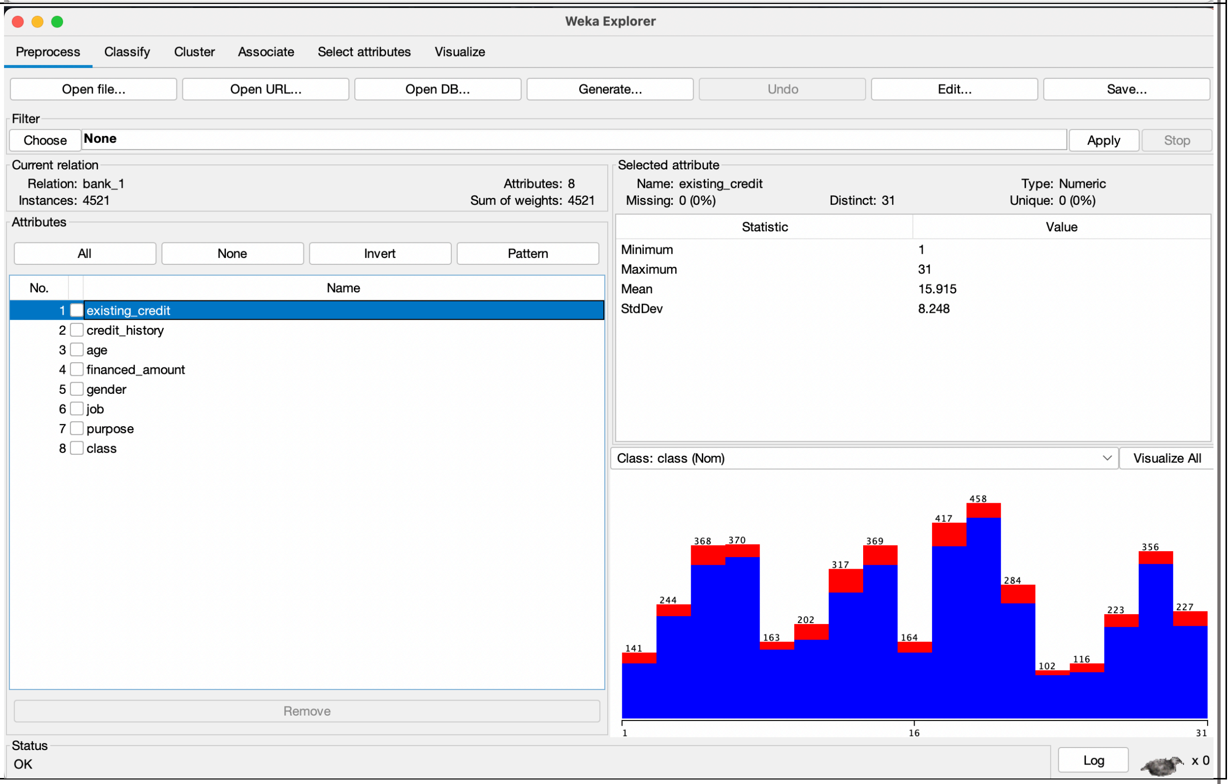
File Edit View

bank_1.csv

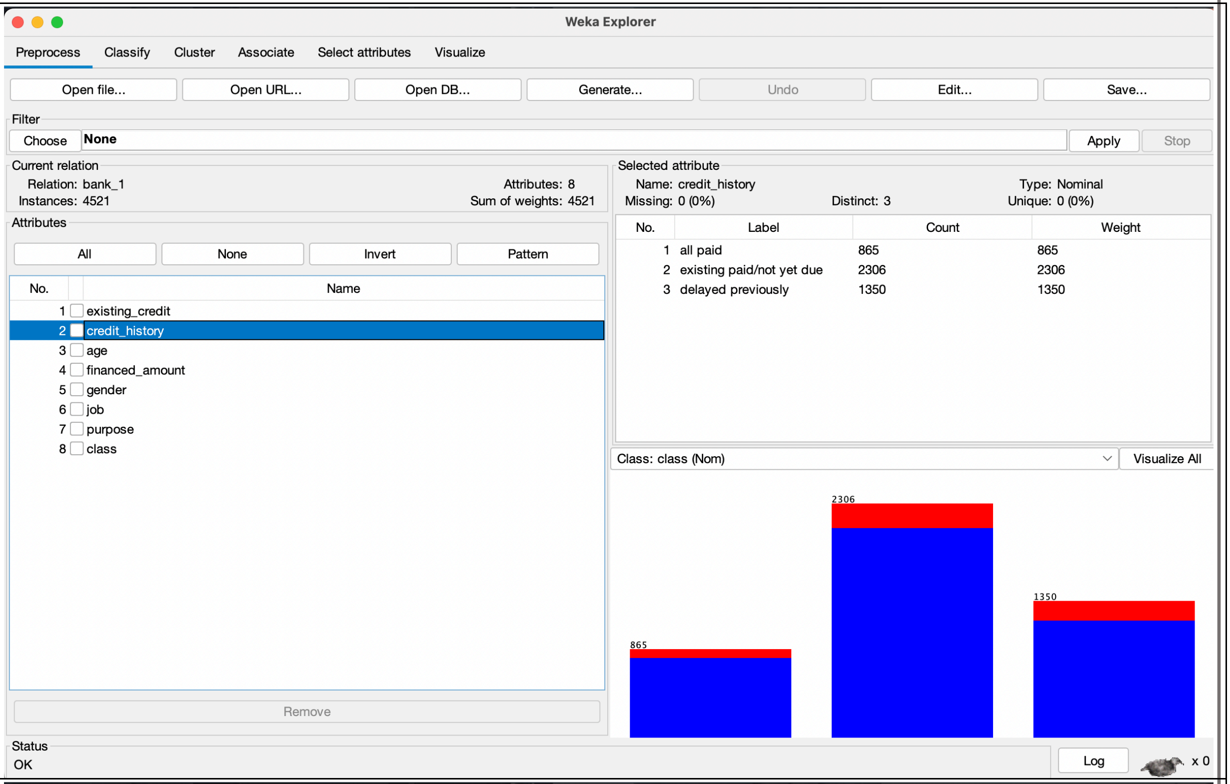
Relation: bank_1

No.	1: existing_credit Numeric	2: credit_history Nominal	3: age Numeric	4: financed_amount Numeric	5: gender Nominal	6: job Nominal	7: purpose Nominal	8: class Nominal
1	19.0	all paid	30.0	1787.0	male	unemployed / retired	Acquisition of land aut...	good
2	11.0	existing paid/not yet due	33.0	4789.0	female	unskilled	Acquisition of seeds, f...	good
3	16.0	delayed previously	35.0	1350.0	female	qualified / self employed	Housing	good
4	3.0	delayed previously	30.0	1476.0	female	qualified / self employed	Housing	good
5	5.0	existing paid/not yet due	59.0	0.0	female	qualified / self employed	Medium Scale Enterpri...	good
6	23.0	delayed previously	35.0	747.0	male	qualified / self employed	Housing	good
7	14.0	delayed previously	36.0	307.0	female	self-employed	Others	good
8	6.0	existing paid/not yet due	39.0	147.0	female	skilled	Construction, Acquisiti...	good
9	14.0	delayed previously	41.0	221.0	female	skilled	Housing	good
10	17.0	all paid	43.0	-88.0	female	unskilled	Acquisition of seeds, f...	good
11	20.0	existing paid/not yet due	39.0	9374.0	female	unskilled	Acquisition of seeds, f...	good
12	17.0	existing paid/not yet due	43.0	264.0	female	qualified / self employed	Small Scale Enterprise.	good
13	13.0	delayed previously	36.0	1109.0	male	skilled	Construction, Acquisiti...	good
14	30.0	existing paid/not yet due	20.0	502.0	male	student	Studies / abroad	bad
15	29.0	existing paid/not yet due	31.0	360.0	female	qualified / self employed	Medium Scale Enterpri...	good
16	29.0	delayed previously	40.0	194.0	male	qualified / self employed	Housing	good
17	27.0	existing paid/not yet due	56.0	4073.0	male	skilled	Construction, Acquisiti...	good
18	20.0	delayed previously	37.0	2317.0	female	qualified / self employed	Small Scale Enterprise.	good
19	23.0	all paid	25.0	-221.0	female	qualified / self employed	Medium Scale Enterpri...	good
20	7.0	existing paid/not yet due	31.0	132.0	male	unskilled	Acquisition of seeds, f...	good
21	18.0	all paid	38.0	0.0	female	qualified / self employed	Housing	good
22	19.0	delayed previously	42.0	16.0	male	qualified / self employed	Housing	good
23	12.0	existing paid/not yet due	44.0	106.0	male	unskilled	Acquisition of seeds, f...	good
24	7.0	existing paid/not yet due	44.0	93.0	male	skilled	Housing	good
25	30.0	delayed previously	26.0	543.0	male	unskilled	Small Scale Enterprise	good
26	20.0	delayed previously	41.0	5883.0	male	qualified / self employed	Housing	good
27	5.0	all paid	55.0	627.0	female	qualified / self employed	Medium Scale Enterpri...	good
28	17.0	all paid	67.0	696.0	male	unemployed / retired	Others	good
29	30.0	existing paid/not yet due	56.0	784.0	male	self-employed	Others	good
30	21.0	existing paid/not yet due	53.0	105.0	male	qualified / self employed	Small Scale Enterprise.	good
31	14.0	existing paid/not yet due	68.0	4189.0	male	unemployed / retired	Others	bad

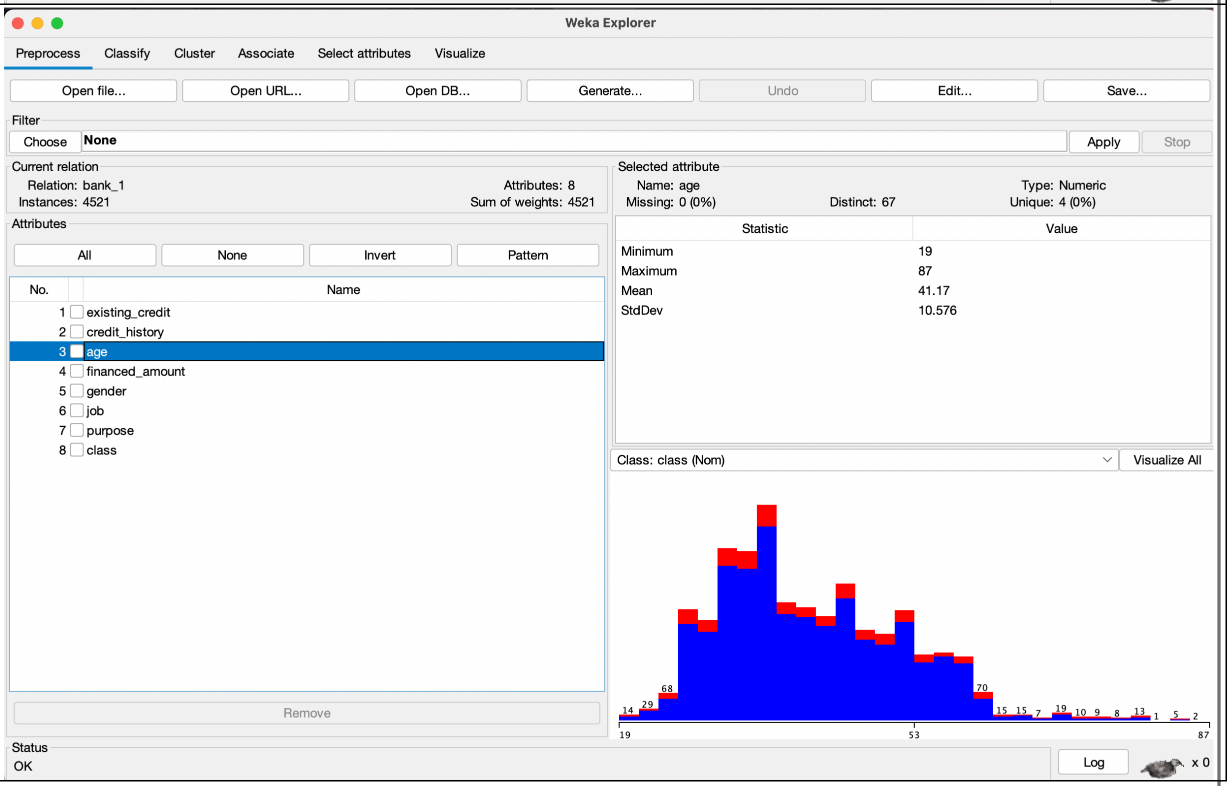
Existing_credit attribute

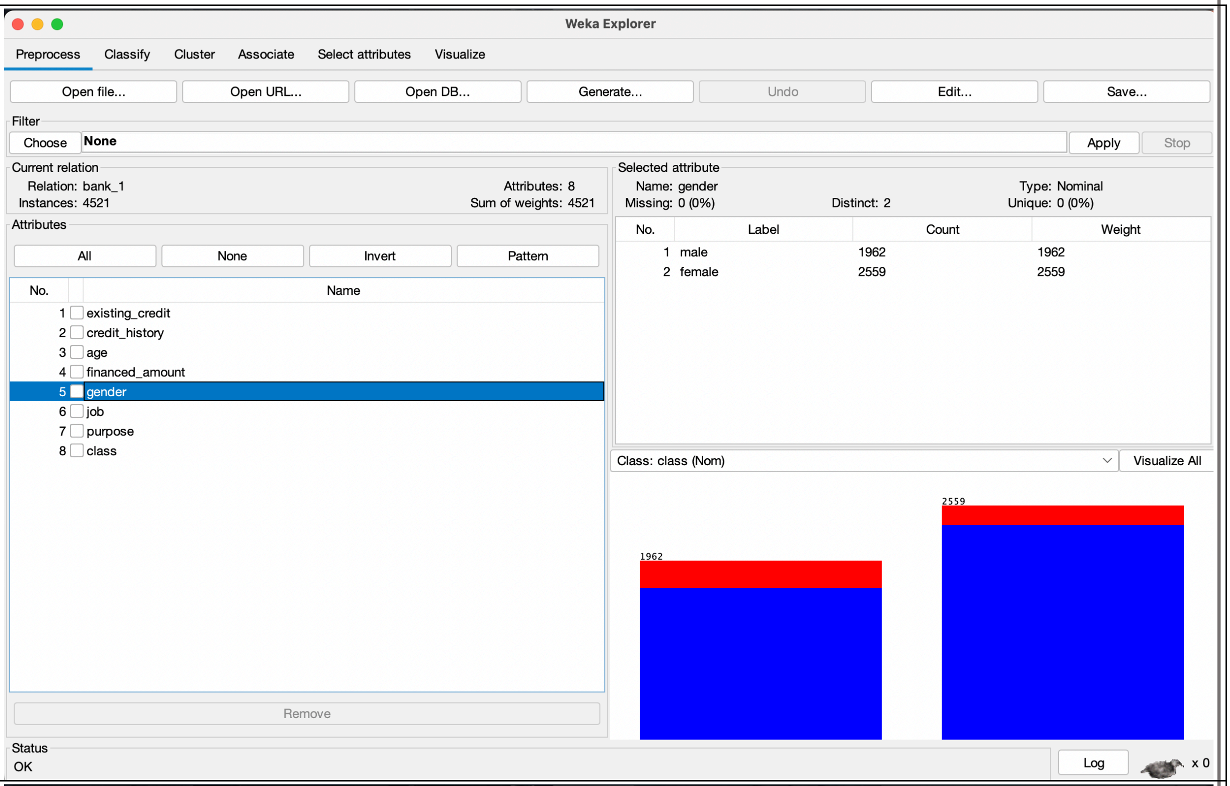
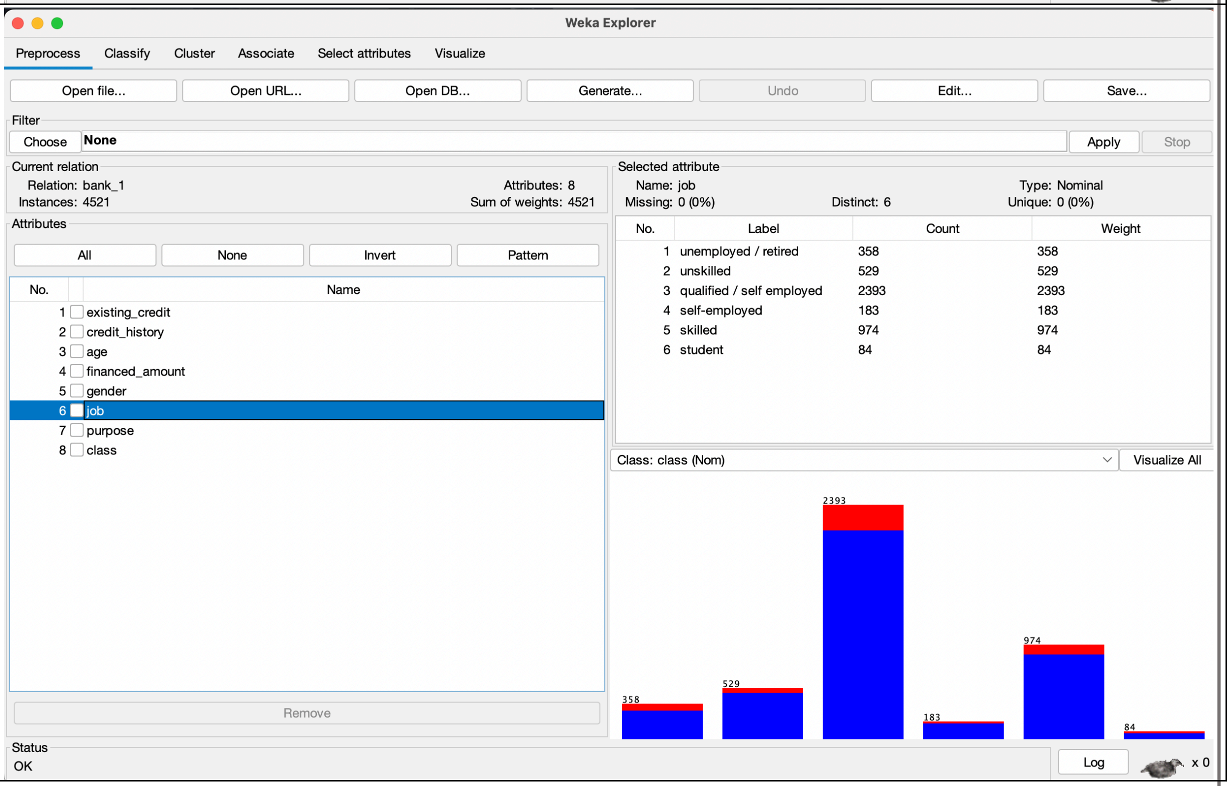


Credit_history attribute

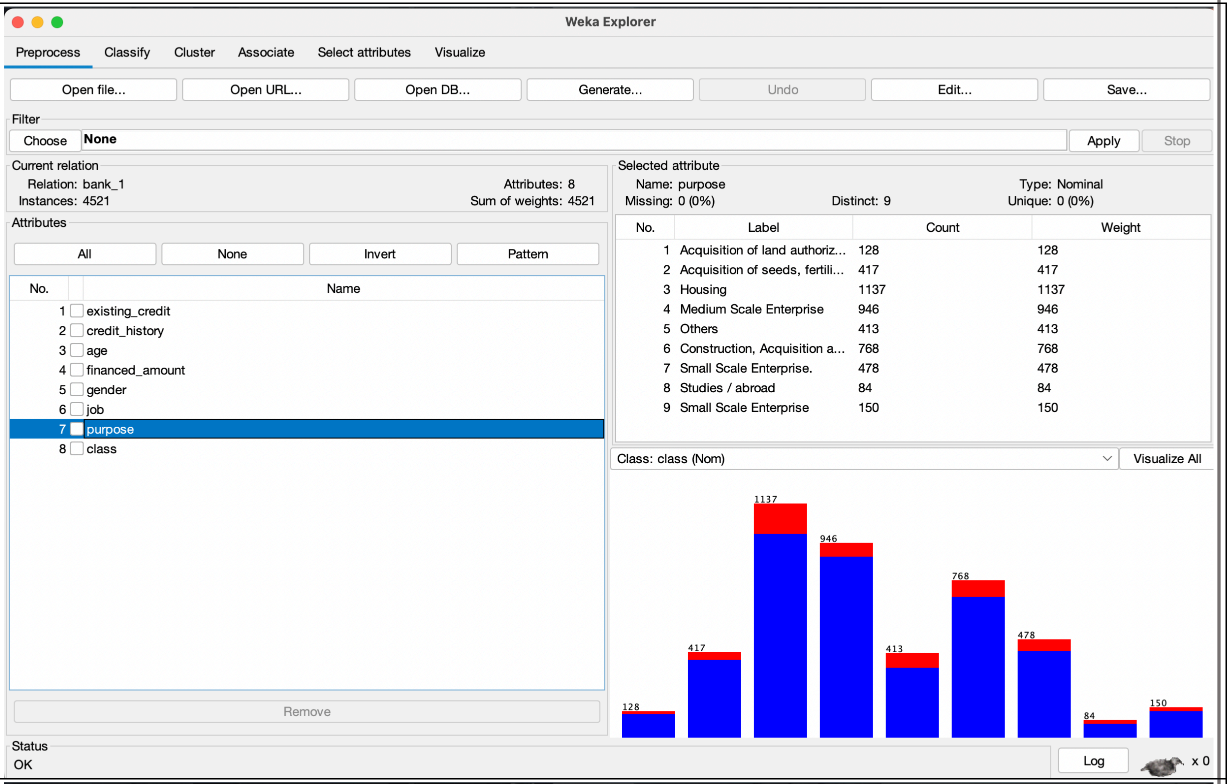


Age attribute

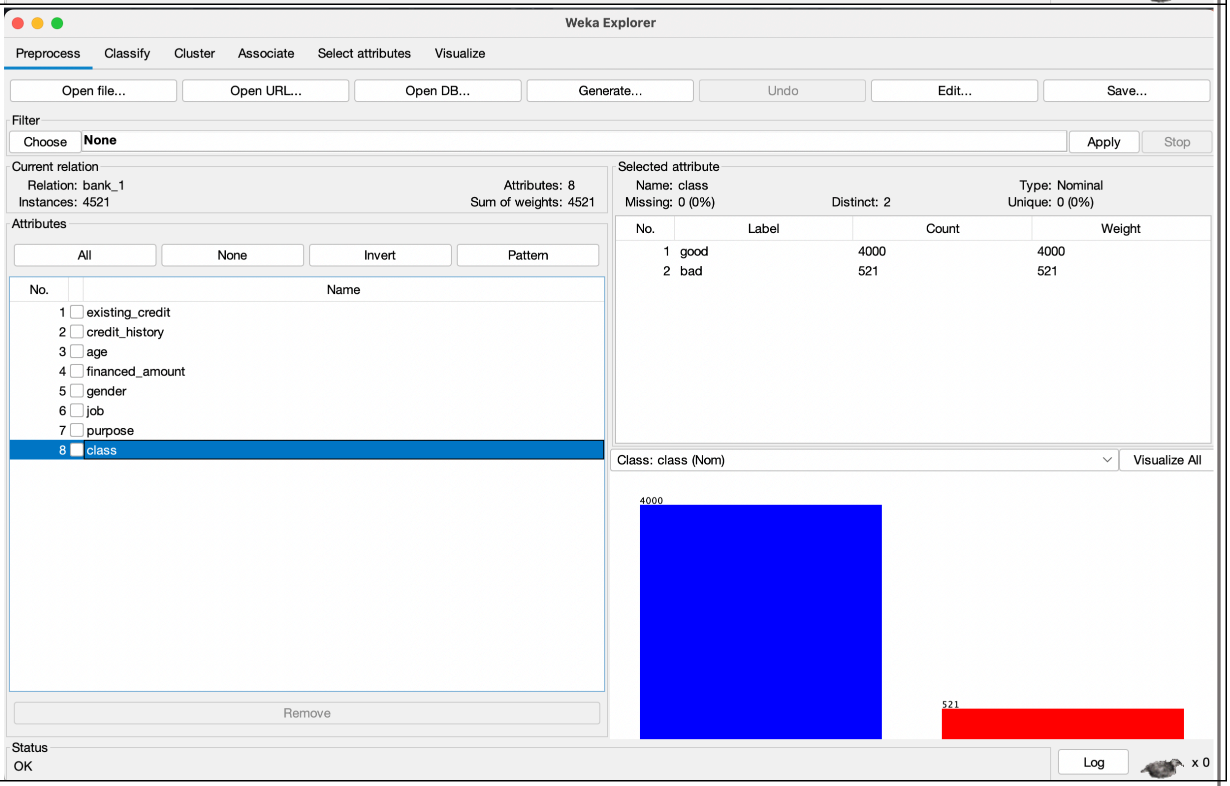


Gender attribute	
Job attribute	

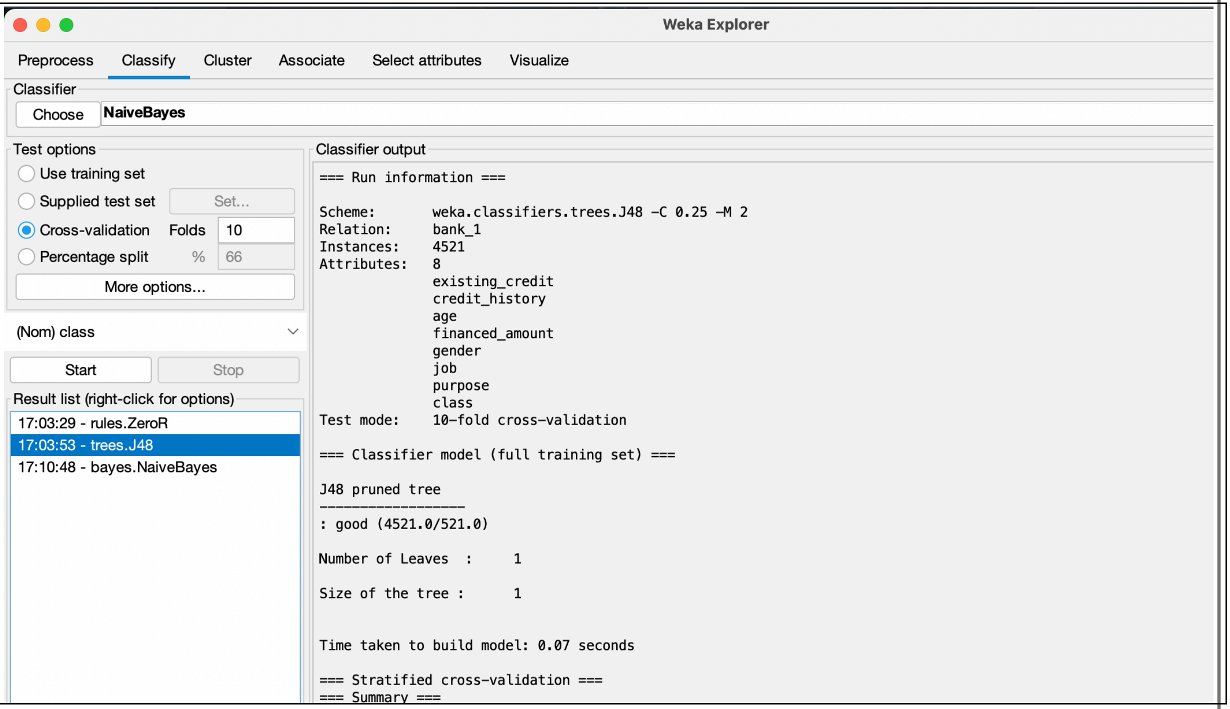
Purpose attribute



Class



J48 classification



After comparing the algorithms, I found that the J48 classification showed better predictability over other algorithms. This helped identify the applicant's risk factor for providing credit on loans.

The processing was done on the dataset of 4521 instances and eight attributes using the 10-fold cross-validation testing mode. The accuracy was 88.37%.
(S., Singaravelan & Arun, R. & Shunmugam, D. & Soundar, K. & Mayakrishnan, R. & Murugan, D.. (2018).)

7 Conclusion

The process of CRISP-DM along with WEKA has been used in this solution to build a model that can predict potential loan applications for customers as good or bad using the customers' records and history.

The data preprocessing and cleaning have also been highlighted (refer to bank_1.csv), showing how it plays a vital role in high accuracy. The results are based on the J48 algorithm giving a curative effect of 88.37% compared to other algorithms.

This process can be enhanced further to identify target loan markets and future income-enhancing action plans, reduce default risk, and improve loan products.

8 References

- Moin, K.I., & Ahmed, Q.B. (2012). Use of Data Mining in Banking.
- S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014

- Hamid, Aboobyda & Ahmed, Tarig. (2016). Developing Prediction Model of Loan Risk in Banks Using Data Mining. Machine Learning and Applications: An International Journal. 3. 1-9. 10.5121/mlaij.2016.3101.
- Akcura, Korhan & Chhibber, Appan. (2018). Design and Comparison of Data Mining Techniques for Predicting
- S., Singaravelan & Arun, R. & Shunmugam, D. & Soundar, K. & Mayakrishnan, R. & Murugan, D.. (2018). ANALYSIS OF CLASSIFICATION ALGORITHMS ON DIFFERENT DATASETS. Review of Innovation and Competitiveness. 4. 41-54. 10.32728/ric.2018.42/3.
- Jothikumar, R. & Sivabalan, R.V. & Sivarajan, E.. (2015). Accuracies of J48 weka classifier with different supervised weka filters for predicting heart diseases. 10. 7788-7793.
- Shivangi Gupta & Neeta Verma. (2016). Comparative Analysis of Classification Algorithms using WEKA tool.International Journal of Scientific & Engineering Research, Volume 7, Issue 8, August-2016. ISSN 2229-5518.